

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/62, 15/15, C07K 14/815, 14/61,</b> <b>C12N 9/50, 15/57, A61K 38/58, 38/27,</b> <b>A23L 1/305</b>	<b>A1</b>	<b>(11) International Publication Number: WO 98/49326</b> <b>(43) International Publication Date: 5 November 1998 (05.11.98)</b>
<b>(21) International Application Number:</b> PCT/CA98/00398 <b>(22) International Filing Date:</b> 23 April 1998 (23.04.98)  <b>(30) Priority Data:</b> 60/044,254 25 April 1997 (25.04.97) US  <b>(71) Applicant (for all designated States except US):</b> SEMBIOSYS GENETICS INC. [CA/CA]; Suite 204, 609-14 Street N.W., Calgary, Alberta T2N 2A1 (CA).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> MOLONEY, Maurice [IE/CA]; 34 Edgebrook Cove, N.W., Calgary, Alberta T3A 5N5 (CA). ALCANTARA, Joenel [CA/CA]; 3 Castledale Place N.E., Calgary, Alberta T3J 1Y4 (CA). VAN ROOIJEN, Gijs [NL/CA]; 3223 Bearspaw Drive N.W., Calgary, Alberta T2L 1T1 (CA).  <b>(74) Agent:</b> BERESKIN & PARR; 40th Floor, 40 King Street West, Toronto, Ontario M5H 3Y2 (CA).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> METHOD FOR CLEAVAGE OF FUSION PROTEINS  <b>(57) Abstract</b>  An improved method for recovering recombinantly produced polypeptides is described. The method involves expressing the recombinant polypeptide as a fusion protein with a pro-peptide. The pro-peptide-polypeptide fusion protein can be cleaved and the recombinant polypeptide released under the appropriate conditions.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**Title:** METHOD FOR CLEAVAGE OF FUSION PROTEINS

**FIELD OF THE INVENTION**

The present invention relates to an improved method for recovering recombinantly produced polypeptides. The method involves expressing the recombinant polypeptide as a fusion protein with a pro-peptide. The pro-peptide-polypeptide fusion protein can be cleaved and the recombinant protein released under the appropriate conditions.

**BACKGROUND OF THE INVENTION**

The preparation of valuable recombinant (genetically engineered) polypeptides, for example pharmaceutical proteins, relies frequently on techniques which involve the production of these polypeptides as fusion or hybrid proteins. These techniques are based upon the preparation of hybrid genes, i. e. genes comprising genetic material encoding the polypeptide of interest linked to genetic material additional to the gene of interest. Production of the fusion polypeptide involves the introduction of the hybrid gene into a biological host cell system, for example yeast cells, which permits the expression and accumulation of the fusion polypeptide. Recovery of the polypeptide of interest involves the performance of a cleavage reaction which results in the separation of the desired polypeptide from the "fusion partner".

Despite the additional steps which are required to produce a protein of interest as a fusion protein, rather than directly in its active form, the production of hybrid proteins has been found to overcome a number of problems. Firstly, overproduced polypeptides can aggregate in the host cell in insoluble fractions known as inclusion bodies. Conversion of this insoluble material involves often slow and complex refolding methods, making protein purification difficult. Secondly, those proteins which are present in soluble form in the cytoplasm often are subject to degradation by host specific enzymes, thus reducing the amounts of active protein that can be recovered. Linking the polypeptide of interest to a fusion partner has been found to limit these problems. Fusion partners known to the prior art include maltose binding protein (Di Guan et al. (1988) Gene 67: 21-30), glutathione-S-transferase (Johnson (1989) Nature 338: 585-587), ubiquitin (Miller et al. (1989) Biotechnology 7: 698-704),  $\beta$ -galactosidase (Goeddel et al. (1979) Proc. Natl. Acad. Sci. (USA) 76: 106-110), and thioredoxin (LaVallie et al. (1993) Biotechnology 11:187-193).

It has also been proposed to employ fusion partners as affinity peptides. This methodology facilitates the isolation and recovery of the fusion peptide from the host cells by exploiting the physico-chemical properties of the fusion partner. (See, for example, WO 91/11454).

Finally, the use of a fusion partner may enable the production of a peptide which would otherwise be too small to accumulate and recover efficiently from a

- 2 -

recombinant host cell system. This technology is described, for example, by Schultz et al., (1987, J. Bacteriol. 169: 5385-5392)

5 All of these procedures result in the production a hybrid protein in which the protein of interest is linked to an additional polypeptide. In order to recover the active polypeptide it is, in general, necessary to separate the fusion partner from the polypeptide of interest. Most commonly, a cleavage reaction, either by enzymatic or by chemical means, is performed. Such reactions employ agents that act by hydrolysis of peptide bonds and the specificity of the cleavage agent is determined by the identity of the amino acid residue at or near the peptide bond which is cleaved.

10 Enzymes known to the prior art as "proteolytic enzymes" have been found to be particularly well suited for the cleavage of fusion proteins. The cleavage reaction is performed by contacting the fusion protein with a proteolytic enzyme under appropriate conditions. An example of this methodology is described in US Patent 4,743,679 which discloses a process for the production of human epidermal growth factor comprising  
15 cleavage of a fusion protein by *Staphylococcus aureus* V8 protease.

By contrast, chemical cleavage involves the use of chemical agents which are known to permit hydrolysis of peptide bonds under specific conditions. Cyanogenbromide, for example, is known to cleave the polypeptide chain at a methionine residue. A hydrolysis reaction for the cyanogenbromide cleavage of the proteins urease and  
20 phosphorylase b based on this technique is described by Sekita et al. ((1975), Keio J. Med. 24: 203-210).

Both chemical and enzymatic cleavage reactions require the presence of a peptide bond which can be cleaved by the cleavage agent which is employed. For this reason it is often desirable to place an appropriate target sequence at the junction of the fusion partner and the target protein. Fusion peptides comprising "linker" sequences  
25 containing a target for a proteolytic enzyme may readily be constructed using conventional art-recognized genetic engineering techniques.

Despite their great utility, the prior art cleavage methods have been recognized to be either inefficient or lack cleavage specificity. Inefficient cleavage results  
30 in low protein purification efficiency, while the lack of cleavage specificity results in cleavage at several locations resulting in product loss and generation of contaminating fragments. This results frequently in the recovery of only a small fraction of the desired protein. In addition, the currently widely used proteolytic enzymes, such as blood clotting factor Xa and thrombin, are expensive, and contamination of final product with blood  
35 pathogens is a consideration.

In view of these shortcomings, the limitations of the cleavage methods known to the prior art are apparent.

- 3 -

Zymogens, such as pepsin and chymosin, are enzymes which are synthesized as inactive precursors *in vivo*. Under appropriate conditions, zymogens are activated to form the mature active protein in a process involving the cleavage of an amino-terminal peptide which can be referred to as the "pro-peptide", "pro-region" or "pro-sequence". Activation of zymogens may require the presence of an additional specific proteolytic enzyme, for example various hormones, such as insulin, are processed by a specific proteolytic enzyme. Alternatively, activation may occur without an additional enzymatic catalyst. These kinds of zymogens are frequently referred to as "autocatalytically maturing" zymogens. Examples of autocatalytically maturing zymogens include pepsin, pepsinogen and chymosin which are activated by an acidic environment, for example in the mammalian stomach.

The autocatalytic activation and processing of zymogens has been documented extensively (see for example, McCaman and Cummings, (1986), J. Biol. Chem. 261: 15345-15348; Koelsch et al. (1994). FEBS Letters 343: 6-10). It has also been documented that activation of the zymogen does not necessarily require a physical linkage of the pro-peptide to the mature protein (Silen et al. (1989), Nature, 341: 462-464).

There is a need for an improved process for recovering recombinantly produced polypeptides from their expression systems.

#### SUMMARY OF THE INVENTION

The present inventors have developed a novel method for recovering recombinantly produced polypeptides. The method involves expressing the polypeptide as a fusion protein with a pro-peptide so that the recombinant polypeptide can be cleaved from the pro-peptide under the appropriate conditions.

In one aspect, the invention provides a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising a first nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen and a second nucleic acid sequence encoding a polypeptide that is heterologous to the pro-peptide.

In another aspect the present invention provides a fusion protein comprising (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is heterologous to the pro-peptide. In one embodiment, the heterologous polypeptide is a therapeutic or nutritional peptide and the fusion protein may be administered as a pharmaceutical or food composition. In such an embodiment the heterologous polypeptide may be cleaved once the composition is delivered to the host as a result of the physiological conditions at the target organ, tissue or in the bodily fluid.

In a further aspect, the present invention provides a method for the preparation of a recombinant polypeptide comprising

(a) introducing into a host cell an expression vector comprising:

- 4 -

- (1) a nucleic acid sequence capable of regulating transcription in a host cell, operatively linked to
  - (2) a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising (a) a nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen, linked in reading frame to (b) a nucleic acid sequence heterologous to the pro-peptide and encoding the recombinant polypeptide; operatively linked to
  - (3) a nucleic acid sequence encoding a termination region functional in the host cell,
- (b) growing the host cell to produce said fusion protein; and
  - (c) altering the environment of the fusion protein so that the pro-peptide is cleaved from the fusion protein to release the recombinant polypeptide.

The environment of the fusion protein can be altered using many means including altering the pH, temperature or salt concentration or other alterations that permit to pro-peptide to self-cleave from the fusion protein to release to recombinant polypeptide. In a preferred embodiment, the mature zymogen is added to the method in step (c) to assist in the cleavage of the propeptide from the fusion protein.

Other features and advantages of the present invention will become readily apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art of this detailed description.

## 25 **BRIEF DESCRIPTION OF THE DRAWINGS**

The invention will now be described in relation to the drawings in which:

Figure 1 is the nucleic acid (SEQ.ID.NO.:1) and deduced amino acid sequence (SEQ.ID.NO.:2) of a GST-Chymosin pro-peptide-Hirudin sequence.

Figure 2 is the nucleic acid (SEQ.ID.NO.:3) and deduced amino acid sequence (SEQ.ID.NO.:4) of a poly histidine tagged chymosin pro-peptide carp growth hormone (His-Pro-cGH) fusion protein.

Figure 3 is a schematic diagram of the Pro-cGH fusion construct.

Figure 4 illustrates the in vitro cleavage of purified His-Pro-cGH.

Figure 5 illustrates the in vivo cleavage of purified His-Pro-cGH.

## 35 **DETAILED DESCRIPTION OF THE INVENTION**

As hereinbefore mentioned, the present invention relates to a novel method for preparing and recovering recombinant polypeptides, chimeric nucleic acid sequences



- 5 -

encoding fusion proteins and fusion proteins useful in pharmaceutical and nutritional compositions.

Accordingly, the present invention provides a method for the preparation of a recombinant polypeptide comprising:

- 5 (a) introducing into a host cell an expression vector comprising:
- (1) a nucleic acid sequence capable of regulating transcription in a host cell, operatively linked to
  - (2) a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising (a) a nucleic acid sequence  
10 encoding a pro-peptide derived from an autocatalytically maturing zymogen, linked in reading frame to (b) a nucleic acid sequence heterologous to the pro-peptide and encoding the recombinant polypeptide, operatively linked to
  - (3) a nucleic acid sequence encoding a termination region functional in  
15 said host cell,
- b) growing the host cell to produce said fusion protein; and
- c) altering the environment of the fusion protein so that the pro-peptide is cleaved from the fusion protein to release the recombinant polypeptide.

The environment of the fusion protein can be altered using many means  
20 including altering the pH, temperature or salt concentration or other alterations that permit to pro-peptide to self-cleave from the fusion protein to release to recombinant polypeptide. In a preferred embodiment, the mature zymogen is added to the method in step (c) to assist in the cleavage of the propeptide from the fusion protein

The term "pro-peptide" as used herein means the amino terminal portion of  
25 a zymogen or a functional portion thereof up to the maturation site.

The term "autocatalytically maturing zymogen" as used herein means that:  
(i) the zymogen can be processed to its active form without requiring an additional specific protease and that (ii) the mature form of the zymogen can assist in the cleavage reaction.

The term "mature zymogen" as used herein means a zymogen that does not  
30 contain the pro-peptide sequence or portion.

The polypeptide can be any polypeptide that is heterologous to the pro-peptide, meaning that it is not the mature protein that is normally associated with the pro-peptide as a zymogen.

In another aspect, the invention provides a chimeric nucleic acid sequence  
35 encoding a fusion protein, the chimeric nucleic acid sequence comprising a first nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen and a second nucleic acid sequence encoding a polypeptide that is heterologous to the pro-peptide.

- 6 -

The chimeric nucleic acid sequence generally does not include a nucleic acid sequence encoding the entire zymogen.

The chimeric nucleic acid sequences which encode the fusion proteins of the present invention can be incorporated in a known manner into a recombinant expression  
5 vector which ensures good expression in a host cell.

Accordingly, the present invention also includes a recombinant expression vector comprising a chimeric nucleic acid molecule of the present invention operatively linked to a regulatory sequence and termination region suitable for expression in a host cell.

The term "nucleic acid sequence" refers to a sequence of nucleotide or  
10 nucleoside monomers consisting of naturally occurring bases, sugars, and intersugar (backbone) linkages. The term also includes modified or substituted sequences comprising non-naturally occurring monomers or portions thereof, which function similarly. The nucleic acid sequences of the present invention may be ribonucleic (RNA) or deoxyribonucleic acids (DNA) and may contain naturally occurring bases including adenine, guanine, cytosine,  
15 thymidine and uracil. The sequences may also contain modified bases such as xanthine, hypoxanthine, 2-aminoadenine, 6-methyl, 2-propyl and other alkyl adenines, 5-halo uracil, 5-halo cytosine, 6-aza uracil, 6-aza cytosine and 6-aza thymine, pseudo uracil, 4-thiouracil, 8-halo adenine, 8-aminoadenine, 8-thiol adenine, 8-thiolalkyl adenines, 8-hydroxyl adenine and other 8-substituted adenines, 8-halo guanines, 8-amino guanine, 8-thiol  
20 guanine, 8-thiolalkyl guanines, 8-hydroxyl guanine and other 8-substituted guanines, other aza and deaza uracils, thymidines, cytosines, adenines, or guanines, 5-trifluoromethyl uracil and 5-trifluoro cytosine.

The term "suitable for expression in a host cell" means that the recombinant expression vectors contain the chimeric nucleic acid sequence of the invention, a regulatory  
25 sequence and a termination region, selected on the basis of the host cells to be used for expression, which is operatively linked to the chimeric nucleic acid sequence. Operatively linked is intended to mean that the chimeric nucleic acid sequence is linked to a regulatory sequence and a termination region in a manner which allows expression of the chimeric sequence. Regulatory sequences and termination regions are art-recognized and are selected  
30 to direct expression of the desired protein in an appropriate host cell. Accordingly, the term regulatory sequence includes promoters, enhancers and other expression control elements. Such regulatory sequences are known to those skilled in the art or one described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990) can be used. It should be understood that the design of the expression  
35 vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Such expression vectors can be used to transform cells to thereby produce fusion proteins or peptides encoded by nucleic acids as described herein.

- 7 -

The recombinant expression vectors of the invention can be designed for expression of encoded fusion proteins in prokaryotic or eukaryotic cells. For example, fusion proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using, for example baculovirus), yeast cells, plant cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). The type of host cell which is selected to express the fusion protein is not critical to the present invention and may be as desired.

Expression in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of the fusion proteins. Inducible expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). While target gene expression relies on host RNA polymerase transcription from the hybrid trp-lac fusion promoter in pTrc, expression of target genes inserted into pET 11d relies on transcription from the T7 gn10-lac 0 fusion promoter mediated by coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$  prophage harboring a T7 gn1 under the transcriptional control of the lacUV 5 promoter. Another attractive bacterial expression system is the pGEX expression system (Pharmacia) in which genes are expressed as fusion products of glutathione-S-transferase (GST), allowing easy purification of the expressed gene from a GST affinity column.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinantly expressed proteins (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the chimeric DNA to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) Nuc. Acids Res. 20: 2111-2118). Such alteration of nucleic acid sequences of the invention could be carried out by standard DNA synthesis techniques.

Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari. et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA).

Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) Virology).

Vectors such as the Ti and Ri plasmids are available for transformation and expression of plants. These vectors specify DNA transfer functions and are used when it is

desired that the constructs are introduced into the plant and stably integrated into the genome via *Agrobacterium*-mediated transformation.

A typical construct consists, in the 5' to 3' direction, of a regulatory region complete with a promoter capable of directing expression in plant, a protein coding region, and a sequence containing a transcriptional termination signal functional in plants. The sequences comprising the construct may be either natural or synthetic or any combination thereof.

Both non-seed specific promoters, such as the 35-S CaMV promoter (Rothstein et al., (1987), Gene 53: 153-161) and, if seed specific expression is desired, seed-specific promoters such as the phaseolin promoter (Sengupta-Gopalan et al., (1985), PNAS USA 82: 3320-3324) or the *Arabidopsis* 18 kDa oleosin (Van Rooijen et al., (1992) Plant Mol. Biol. 18: 1177-1179) promoters may be used. In addition to the promoter, the regulatory region contains a ribosome binding site enabling translation of the transcripts in plants and may also contain one or more enhancer sequences, such as the AMV leader (Jobling and Gehrke, (1987), Nature 325: 622-625), to increase the expression of product.

The coding region of the construct will typically be comprised of sequences encoding a pro-peptide region fused in frame to a desired protein and ending with a translational termination codon. The sequence may also include introns.

The region containing the transcriptional termination signal may comprise any such sequence functional in plants such as the nopaline synthase termination sequence and additionally may include enhancer sequences to increase the expression of product.

The various components of the construct are ligated together using conventional methods, typically into a pUC-based vector. This construct may then be introduced into an *Agrobacterium* vector and subsequently into host plants, using one of the transformation procedures outlined below.

The expression vectors will normally also contain a marker which enables expression in plant cells. Conveniently, the marker may be a resistance to a herbicide, for example glyphosate, or an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol or the like. The particular marker employed will be one which will permit selection of transformed cells from cells lacking the introduced recombinant nucleic acid molecule.

A variety of techniques is available for the introduction of nucleic acid sequences, in particular DNA into plant host cells. For example, the chimeric DNA constructs may be introduced into host cells obtained from dicotyledonous plants, such as tobacco, and oleaginous species, such as *B. napus* using standard *Agrobacterium* vectors; by a transformation protocol such as that described by Moloney et al., (1989), (Plant Cell Rep., 8: 238-242) or Hinchee et al., (1988), (Bio/Technol., 6: 915-922); or other techniques known to those skilled in the art. For example, the use of T-DNA for transformation of plant cells

has received extensive study and is amply described in EPA Serial No. 120,516; Hoekema et al., (1985), (Chapter V, In: *The Binary Plant Vector System* Offset-drukkerij Kanters B.V., Alblasterdam); Knauf, et al., (1983), (*Genetic Analysis of Host Range Expression by Agrobacterium*, p. 245, In *Molecular Genetics of the Bacteria-Plant Interaction*, Puhler, A. ed., Springer-Verlag, NY); and An et al., (1985), (*EMBO J.*, 4: 277-284). Conveniently, explants may be cultivated with *A. tumefaciens* or *A. rhizogenes* to allow for transfer of the transcription construct to the plant cells. Following transformation using *Agrobacterium* the plant cells are dispersed in an appropriate medium for selection, subsequently callus, shoots and eventually plantlets are recovered. The *Agrobacterium* host will harbour a plasmid comprising the vir genes necessary for transfer of the T-DNA to the plant cells. For injection and electroporation, (see below) disarmed Ti-plasmids (lacking the tumour genes, particularly the T-DNA region) may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of the constructs described herein to obtain transformation and expression in a wide variety of monocotyledonous and dicotyledonous plants and other organisms. These techniques are especially useful for species that are intractable in an *Agrobacterium* transformation system. Other techniques for gene transfer include biolistics (Sanford, (1988), *Trends in Biotech.*, 6: 299-302), electroporation (Fromm et al., (1985), *Proc. Natl. Acad. Sci. USA*, 82: 5824-5828; Riggs and Bates, (1986), *Proc. Natl. Acad. Sci. USA* 83: 5602-5606) or PEG-mediated DNA uptake (Potrykus et al., (1985), *Mol. Gen. Genet.*, 199: 169-177).

In a specific application, such as to *B. napus*, the host cells targeted to receive recombinant DNA constructs typically will be derived from cotyledonary petioles as described by Moloney et al., (1989, *Plant Cell Rep.*, 8: 238-242). Other examples using commercial oil seeds include cotyledon transformation in soybean explants (Hinchee et al., (1988), *Bio/Technology*, 6: 915-922) and stem transformation of cotton (Umbeck et al., (1981), *Bio/Technology*, 5: 263-266).

Following transformation, the cells, for example as leaf discs, are grown in selective medium. Once shoots begin to emerge, they are excised and placed onto rooting medium. After sufficient roots have formed, the plants are transferred to soil. Putative transformed plants are then tested for presence of a marker. Southern blotting is performed on genomic DNA using an appropriate probe, for example a chymosin pro-sequence, to show that integration of the desired sequences into the host cell genome has occurred.

Transformed plants grown in accordance with conventional ways, are allowed to set seed. See, for example, McCormick et al. (1986, *Plant Cell Reports*, 5: 81-84). Northern blotting can be carried out using an appropriate gene probe with RNA isolated from tissue in which transcription is expected to occur, such as a seed embryo. The size of the transcripts can then be compared with the predicted size for the fusion protein transcript.

Two or more generations of transgenic plants may be grown and either crossed or selfed to allow identification of plants and strains with desired phenotypic characteristics including production of recombinant proteins. It may be desirable to ensure homozygosity of the plants, strains or lines producing recombinant proteins to assure continued inheritance of the recombinant trait. Methods of selecting homozygous plants are well known to those skilled in the art of plant breeding and include recurrent selfing and selection and anther and microspore culture. Homozygous plants may also be obtained by transformation of haploid cells or tissues followed by regeneration of haploid plantlets subsequently converted to diploid plants by any number of known means, (e.g.: treatment with colchicine or other microtubule disrupting agents).

The polypeptide of the present invention may be any polypeptide that is not normally fused to the pro-peptide used in the method. The polypeptide is preferentially stable under cleavage conditions, for example at acidic pH, and the polypeptide may be activated after cleavage upon adjusting the pH, or altering the environment otherwise so that conditions optimal for enzymatic activity are generated. The cleavage reaction may be performed any time upon commencement of the production of the fusion protein in a recombinant cell system. In preferred embodiments the cleavage reaction is performed using crude cellular extracts producing the recombinant protein or any purified fraction thereof.

The pro-peptide used in the present invention may be any pro-peptide derived from any autocatalytically maturing zymogen, including those pro-peptides derived from proteases, including aspartic proteases, serine proteases and cysteine proteases. In preferred embodiments of the invention, the pro-peptide is derived from chymosin, pepsin, HIV-1 protease, pepsinogen, cathepsin or yeast proteinase A. The amino acid and/or DNA sequences of pepsinogen (Ong et al. (1968), J. Biol. Chem. 6104-6109; Pedersen et al., (1973), FEBS Letters, 35: 255-526), chymosin (Foltmann et al., (1977); Harris et al., (1982), Nucl. Acids. Res., 10: 2177-2187), yeast proteinase A (Ammerer et al., (1986), Mol. Cell. Biol. 6: 2490-2499; Woolford et al., (1986), Mol. Cell. Biol. 6: 2500-2510), HIV-1 protease (Ratner et al., (1987), AIDS Res. Human Retrovir. 3: 57-69.), cathepsin (McIntyre et al., (1994), J. Biol. Chem. 269: 567-572) and pepsin are available (Koelsch et al. (1994), FEBS Lett. 343: 6-10). Based on these sequences cDNA clones comprising the genetic material coding for the pro-peptides may be prepared and fusion genes may be prepared in accordance with the present invention and practising techniques commonly known to those skilled in the art (see e.g. Sambrook et al. (1990), Molecular Cloning, 2nd Ed., Cold Spring Harbor Press).

To identify other pro-sequences having the desired characteristics, where a zymogen undergoing autocatalytic cleavage has been isolated (for example chymosin and yeast protein A), the protein may be partially sequenced, so that a nucleic acid probe may

be designed to identify other pro-peptides. The nucleic acid probe may be used to screen cDNA or genomic libraries prepared from any living cell or virus. Sequences which hybridize with the probe under stringent conditions may then be isolated.

Other pro-sequences may also be isolated by screening of cDNA expression  
5 libraries. Antibodies against existing pro-peptides may be obtained and cDNA expression libraries may be screened with these antibodies essentially as described by Huynh et al. (1985, in DNA cloning, Vol. 1, a Practical Approach, ed. D. M. Glover, IRL Press). Expression libraries may be prepared from any living cell or virus.

Other zymogens which are autocatalytically processed may be discovered  
10 by those skilled in the art. The actual pro-sequence which is selected is not of critical importance and may be as desired. It is to be clearly understood that the pro-sequence of any autocatalytically maturing zymogen may be employed without departing from the spirit or scope of the present invention.

Upon isolation of a pro-sequence, the pro-peptide encoding genetic material  
15 may be fused to the genetic material encoding polypeptide of interest using DNA cloning techniques known to skilled artisans such as restriction digestion, ligation, gel electrophoresis, DNA sequencing and PCR. A wide variety of cloning vectors are available to perform the necessary cloning steps. Especially suitable for this purpose are the cloning vectors which include a replication system that is functional in *E. coli* such as pBR322, the  
20 pUC series, M13mp series, pACYC184, pBluescript etc. Sequences may be introduced into these vectors and the vectors may be used to transform the *E. coli* host, which may be grown in an appropriate medium. Plasmids may be recovered from the cells upon harvesting and lysing the cells.

The invention also includes the full length pro-peptide as well as  
25 functional portions of the pro-peptide or functional mutated forms of the pro-peptide. Mutated forms of the pro-peptide may be used to obtain specific cleavage between the pro-peptide and a heterologous protein. Mutations in the pro-peptide could alter the optimal conditions, such as temperature, pH and salt concentration, under which cleavage of a heterologous peptide is achieved (McCaman, M.T. and Cummings, D.B., (1986), J. Biol.  
30 Chem. 261:15345-15348). Depending on the pro-peptide, cleavage of the heterologous protein from various pro-peptides, will be optimal under varying different conditions. Thus the invention will be amenable to heterologous proteins which are preferentially cleaved under a variety of desirable conditions.

The nucleic acid sequence encoding the heterologous polypeptide may be  
35 fused upstream or downstream of the nucleic acid sequence encoding the pro-peptide and concatamers containing repetitive units of the pro-peptide fused to the heterologous protein may be employed. In preferred embodiments, the heterologous protein is fused downstream

of the pro-peptide. The nucleic acid sequence encoding the pro-peptide generally does not include the mature form of the zymogen.

In one embodiment, the pro-peptide is a pro-peptide derived from chymosin and the heterologous polypeptide is hirudin (Dodt et al., (1984), FEBS Letters 65:180-183).

- 5 In particular, the present inventors have constructed a chimeric DNA sequence in which the DNA encoding the chymosin pro-peptide was fused upstream of the DNA sequence encoding the leech anticoagulant protein hirudin. The gene fusion (Pro-Hirudin) was expressed in *E. coli* cells. It was found that upon lowering of the pH to pH 2, and more preferably to pH 4.5, and in the presence of a small quantity of mature chymosin, the heterologously fused  
10 protein, hirudin, was efficiently cleaved from the chymosin pro-peptide.

- Autocatalytic cleavage requires an alteration of the environment of the fusion peptide. This may include alterations in pH, temperature, salt concentrations, the concentrations of other chemical agents or any other alteration resulting in environmental conditions that will permit autocatalytic cleavage of the fusion protein. The environment  
15 may be altered by the delivery of the fusion protein into an appropriate cleavage environment. The cleavage environment may be a physiological environment, such as for example in the mammalian stomach, gut, kidneys, milk or blood, or the environmental conditions may be man-made. The cleavage environment may also be generated by the addition of an agent or agents or by altering the temperature of the environment of the  
20 fusion protein. The cleavage reaction may take place when the fusion protein is pure or substantially pure, as well as when it is present in cruder preparations, such as cellular extracts.

- In a preferred embodiment, the inventors have employed mature chymosin to assist in the cleavage reaction. Generally, the addition of the mature enzyme will assist  
25 in the cleavage reaction. The enzyme used for this reaction may be homologous to the pro-peptide, for example, chymosin may be used to assist cleavage of pro-chymosin fused to a desired protein, or heterologous to the pro-peptide, for example, pepsin may be used to assist in cleavage of a pro-chymosin fused to a desired protein.

- Although in a preferred embodiment mature chymosin is added, it is  
30 conceivable that the use of other pro-peptides may not require the addition of the mature peptide in order to accomplish efficient cleavage.

- Activation of the fusion protein may be *in vitro* or *in vivo*. In one embodiment, the pro-peptide is used to facilitate cleavage from proteins recombinantly produced on oil bodies as disclosed in PCT application Publication No. WO 96/21029. In  
35 this embodiment, the pro-peptide would be fused downstream of an oil body protein and upstream of the recombinant protein or peptide of interest.

In another *in vivo* application, two vectors would be introduced in the same host. In one vector expression of the zymogen or the mature protein would be controlled by



an inducible promoter system. The other vector would comprise a pro-peptide fused upstream of an heterologous protein of interest. Thus it is possible to control the moment of cleavage of the peptide or protein downstream of the pro-peptide through the promoter which controls expression of the zymogen or the mature protein. Alternatively, the two  
5 expressed genes would be combined in the same vector. In preferred embodiments of this application, the pro-peptide employed is cleaved under physiological conditions.

In another aspect the present invention provides a fusion protein comprising (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is heterologous to the pro-peptide. In one embodiment, the polypeptide  
10 is a therapeutic or nutritional peptide or protein which can be administered as an inactive fusion protein. Activation or maturation through cleavage would only occur upon its delivery at the unique physiological conditions prevalent at the target organ, tissue or bodily fluid for example in the mammalian stomach, gut, kidneys, milk or blood. Cleavage might be enhanced by a protease specific for the peptide, preferably the mature zymogen  
15 homologous to the pro-peptide is used. This method is particularly useful for the delivery of orally ingested vaccines, cytokines, gastric lipase, peptide antibiotics, lactase and cattle feed enzymes which facilitate digestion, such as xylanase and cellulase. For example, a therapeutic or nutritional peptide or protein fused downstream of the chymosin pro-peptide might be activated in the mammal stomach upon ingestion. The mature form of  
20 chymosin or the inactive precursor form of chymosin may be added to assist in the cleavage of the nutritional or therapeutic peptide.

Accordingly, in one embodiment the present invention provides a pharmaceutical composition comprising a fusion protein which comprises (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is  
25 heterologous to the pro-peptide in admixture with a suitable diluent or carrier. The composition may be administered orally, intravenously or via any other delivery route.

The fusion protein and/or mature protein may also be produced in an edible food source, such as animal milk or in an edible crop, which may be consumed without a need for further purification. Accordingly, in another embodiment the present invention  
30 provides a food composition comprising a fusion protein which comprises (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is heterologous to the pro-peptide in admixture with a suitable diluent or carrier. The nutritional composition may be mixed with any liquid or solid food and consumed by a human or animal.

35 The compositions of the invention may include the chimeric nucleic acid sequences or an expression vector containing the chimeric nucleic acid sequences of the present invention. In such an embodiment, the fusion protein is produced *in vivo* in the host animal. The chimeric nucleic acid sequences of the invention may be directly introduced

into cells or tissues *in vivo* using delivery vehicles such as retroviral vectors, adenoviral vectors and DNA virus vectors. The chimeric nucleic acid sequences may also be introduced into cells *in vitro* using physical techniques such as microinjection and electroporation or chemical methods such as co-precipitation and incorporation of nucleic acid into liposomes.

5 Expression vectors may also be delivered in the form of an aerosol or by lavage.

The present invention is also useful in the purification process of recombinant proteins. In one embodiment, a cell extract containing an expressed pro-peptide-heterologous fusion protein is applied to a chromatographic column. Selective binding of the fusion protein to antibodies raised against the pro-peptide sequence and  
10 immobilized onto the column, results in selective retention of the fusion protein. Instead of relying on antibodies against the pro-peptide sequence, a gene encoding another immunogenic domain or a gene encoding a peptide with affinity for a commonly used column material, such as cellulose, glutathione-S-transferase or chitin, or any other desirable tag, may be included in the gene fusion.

15 In another envisaged application, a peptide encoding a sequence which results in anchoring of the fusion protein in the cell wall would be included in the construct. Suitable anchoring proteins for this application would be yeast  $\alpha$ -gluttenin FLO1, the Major Cell Wall Protein of lower eukaryotes, and a proteinase of lactic acid bacteria (PCT 94/18330). Expression of a fusion protein would result in immobilization of the protein of  
20 interest to cell wall. The protein of interest could be isolated by washing the cells with water or washing buffer. Upon cleavage the cells could be removed using a simple centrifugation step and the protein could be isolated from the washing buffer.

The following non-limiting examples are illustrative of the present invention.

## 25 EXAMPLES

### EXAMPLE 1

In the first example, the protein hirudin was prepared as a fusion protein with the chymosin pro-peptide and hirudin was shown to be active in cellular extracts of *E. coli* upon performance of a cleavage reaction.

#### 30 Construction of a pGEX-Pro-Hirudin fusion.

The fusion protein that we studied comprises the pro-peptide of calf chymosin B (Foltmann et al, 1977; Harris et al., 1982, Nucl. Acids. Res., 10: 2177-2187) fused to hirudin variant 1 (Dodt et al., 1984, FEBS Letters 65: 180-183). The hybrid gene which encoded this fusion protein was constructed using standard PCR methods (Horton et al.,  
35 1989, Gene, 77: 61-68). The DNA sequence for this Pro-Hirudin fusion was cloned into pGEX-4T-3 (Pharmacia), downstream of the gene encoding glutathione-S-transferase (GST). The complete sequence of the GST-Pro-Hirudin sequence is shown in Figure 1.

**Growth of *E. coli* transformed with pGEX-4T-3 and pGEX-Pro-Hirudin.**

Plasmids pGEX-4T-3 and pGEX-Pro-Hirudin were transformed into *E. coli* strain DH5 $\alpha$  to allow for high level of expression. A single colony was used to inoculate 5ml LB-amp broth. These cultures were grown overnight. One ml of each overnight culture  
5 was used to inoculate 50 ml of LB-amp broth. These cultures were grown until OD<sub>600</sub> = 0.6. At this OD, IPTG (final concentration 1mM) was added to induce the expression of the GST and GST-Pro-Hirudin fusion proteins. After this induction, the cultures were grown for an additional 3 hours at 37°C. The cells were pelleted at 5000 x g for 10 minutes, and resuspended in 5 ml Tris Buffered Saline (TBS). The resuspended cells were sonicated and  
10 centrifuged at 12000 x g for 15 minutes to separate the inclusion bodies (pellet fraction) from the soluble proteins (supernatant fraction). Western blotting of both the pellet and supernatant fraction indicated that under the growing conditions described above, significant amounts (5-10%) of the GST and GST-Pro-Hirudin protein were found in the supernatant fraction. The rest (90-95%) accumulated in inclusion bodies (results not shown).

**15 Hirudin activity measurements**

The supernatant fractions of both the GST and GST-Pro-Hirudin were tested for anti-thrombin activity. The samples were treated as follows: A) 20  $\mu$ l supernatant + 20  $\mu$ l water B) 20  $\mu$ l supernatant + 20  $\mu$ l of 100 mM Sodium Phosphate pH 2.0 C) 20  $\mu$ l supernatant + 20  $\mu$ l of 100 mM Sodium Phosphate pH 2.0 + 2  $\mu$ g chymosin (Sigma) D) 20  $\mu$ l  
20 supernatant + 20  $\mu$ l of 100 mM Sodium Phosphate pH 4.5 E) 20  $\mu$ l supernatant + 20  $\mu$ l of 100 mM Sodium Phosphate pH 4.5 + 2  $\mu$ g chymosin. These samples were incubated at room temperature for 1 hour. A total of 10  $\mu$ l of the samples was added to 1 ml assay buffer (20 mM Tris [pH 7.5], 100 mM NaCl, 5 mM CaCl<sub>2</sub>, 0.1 unit of thrombin) and incubated for 2-3 minutes before the addition of 50 $\mu$ l p-tos-gly-pro-arg-nitroanilide (1 mM). Thrombin  
25 activity was measured as a function of chromozyme cleavage by monitoring the increase in absorption at 405 nm over time (Chang, 1983, FEBS Letters, 164: 307-313). The  $\Delta_{Abs}$  (405nm) was determined after 2 minutes. The result of the activity measurements are indicated in Table 1.

As can be seen from Table 1, the only extract which exhibited significant  
30 anti-thrombin activity was the extract containing the GST-Pro-Hirudin fusion which was treated at pH 4.5 and supplemented with 2  $\mu$ g chymosin (E). Western blotting (results not shown) indicated that apart from treatment at pH 4.5, complete cleavage was also observed when the GST-Pro-Hirudin fusion which was treated at pH 2.0 and supplemented with 2  $\mu$ g chymosin. It has been well documented that unprocessed chymosin when exposed  
35 at pH 2.0, forms a pseudochymosin, before it matures into chymosin (Foltmann et al., 1977, Scand. J. Clin. Lab. Invest. 42: 65-79; Foltmann, 1992, Proc. Natl. Acad. Sci. 74: 2321-2324; McCaman and Cummings, 1988, J. Biol. Chem. 261: 15345-15348) The pseudo chymosin

- 16 -

cleavage site is located between the Phe<sup>27</sup>-Leu<sup>28</sup> peptide bond and is indicated in Figure 1. The inability of the GST-Pro-Hirudin fusion, which was treated at pH 2.0 and supplemented with 2 µg chymosin, to inhibit thrombin activity might be explained by the fact that cleavage occurred at the Phe<sup>27</sup>-Leu<sup>28</sup> peptide bond rather than at the Phe<sup>43</sup>-Val<sup>44</sup> peptide bond which separates the chymosin pro-peptide from the mature hirudin. It has been well documented that (Loison et al., 1988, Bio/Technology, 6: 72-77) mature hirudin is only active when it does not have any additional amino acids attached to its native N-terminal sequence.

## EXAMPLE 2

In the second example, the protein carp growth hormone (cGH) was prepared as a fusion of pro-chymosin. Carp growth hormone was shown to be present in cellular extracts of *E. coli* upon performance of the cleavage reaction.

### Construction of a pHis-Pro-cGH fusion

A fusion protein was constructed which comprises the pro-peptide of calf chymosin B (Foltmann et al., (1977), Harris et al., 1982, Nucl. Acids Res. 10: 2177-2187 fused to carp growth hormone (Koren et al. (1989), Gene 67: 309-315). The hybrid gene which encoded this fusion protein was constructed using PCR mediated gene-fusion. The DNA sequence for this Pro-cGH fusion was cloned into pUC19 yielding plasmid pPro-cGH. The Pro-cGH gene fusion was released from pPro-cGH by Swal/KpnI digestion and inserted into the PvuII/KpnI site of pRSETB (Invitrogen Corp.), containing a poly-histidine tag, facilitating purification, and an enterokinase recognition and cleavage site to generate pHis-Pro-cGH. The complete sequence of the His-Pro-cGH insert is shown in figure 2.

### Growth of *E. coli* transformed with pHis-Pro-cGH

Plasmid pHis-Pro-cGH was transformed into *E. coli* BL21 strain to allow for high levels of expression. A single colony was used to inoculate LB-amp broth. These cultures were grown overnight. One ml of each o/n culture was used to inoculate 50 ml of LB-amp broth. These cultures were grown until OD<sub>600</sub> = 0.6. At this OD, IPTG (final concentration 0.5 mM) was added to induce the expression of the His-Pro-cGH fusion protein. After this induction, the cultures were grown for an additional 3 hours at 37°C. The cells were pelleted at 5000 x g for 10 minutes, and resuspended in 5 ml PBS (pH 7.3) buffer. The resuspended cells were disrupted by a French-Press and centrifuged at 10,000 x g for 10 minutes. Inclusion bodies were resuspended in 5 ml of water and dissolved by slow addition of NaOH. 1ml of 10 x PBS was added to this solution and the volume was adjusted to 10 ml. The pH of the solution was adjusted to 8.0 by slow addition of HCl and the solution was incubated at 4°C for 2 hours. The pH was adjusted to 7.5 and at this point the solution was centrifuged at 10,000g for 15 minutes to remove insolubles. The fusion protein was then purified by chelating affinity chromatography using Hi-Trap metal binding

columns (Pharmacia). The column was saturated with  $Zn^{++}$  ions and then used to affinity purify His-Pro-cGH fusion protein in accordance with the instructions provided by the manufacturer.

#### Cleavage of cGH produced in *E. coli* transformed with pHis-Pro-cGH

5                   In order to cleave the fusion protein 15  $\mu$ l (ca 1  $\mu$ g) of the protein prep was treated with either 17  $\mu$ l of PBS (Uncut), 14  $\mu$ l of PBS and 3  $\mu$ l of enterokinase (Cut (EK)), or 16  $\mu$ l of phosphate buffer (pH 2) and 1  $\mu$ l of chymosin (Cut (PRO)). All samples were incubated at 37°C for 2 hours and then analysed by SDS-PAGE followed by western blotting. The primary antibody used was a rabbit anti-serum prepared against cGH. The secondary  
10 antibody was goat anti-rabbit IgG which was conjugated with alkaline phosphatase.

As can be seen from figure 3, cleavage of the fusion protein was observed with enterokinase yielding a protein band corresponding to the calculated molecular mass of the Pro-cGH fusion (26 kDa). Similarly the cleavage with chymosin yielded a protein band corresponding to the expected theoretical molecular mass of the cGH (approximately  
15 22 kDa) polypeptide.

#### EXAMPLE 3

In this example, the protein carp growth hormone (cGH) was prepared as a fusion of pro-chymosin. The carp growth hormone fusion protein was cleaved with the gut extract from red turnip beetle, thus illustrating an in vivo application of the invention.

20                   His-Pro-cGH was prepared following the protocol of example 2. Gut extract was prepared from larvae of the red turnip beetle as follows. Red turnip beetle eggs (*Entomoscelis americana* Brown (Coleoptera: Chrysomelidae), were laid by laboratory-reared adults and stored at -20° C for at least three months before use. Eggs were hatched in dishes containing moist filter paper, and larvae were maintained on canola seedlings. Only  
25 larvae that were actively feeding were used. Midguts from second instar larvae were removed by dissection in saline solution and stored in saline at -20° C. Guts were thawed, rinsed in ddH<sub>2</sub>O (50  $\mu$ l per gut). The homogenate was centrifuged at 16,000 xg (10 min, 4° C) and the decanted supernatant was used in the proteolytic assay.

As can be observed in Figure 4, extracts prepared from the gut of red turnip  
30 beetle cleaved the fusion protein and released the cGH polypeptide. Cleavage was not observed to be complete. This could be due to the fact that the pH in the gut extract was not optimal for the cleavage reaction to proceed.

While the present invention has been described with reference to what are presently considered to be the preferred examples, it is to be understood that the invention  
35 is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

All publications, patents and patent applications are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

**DETAILED FIGURE LEGENDS**

Figure 1. The nucleic acid and deduced amino acid sequence of a GST-Pro-Hirudin sequence. The deduced sequence of the chymosin pro-peptide has been underlined and the deduced hirudin protein sequence has been italicized. The hirudin nucleic acid sequence was optimized for plant codon usage. The pseudochymosin cleavage site between Phe27-Leu28 and the peptide bond separating the pro-chymosin and mature hirudin (Phe 42- Val43) are indicated with an arrow (↑).

Figure 2. The nucleic acid and deduced amino acid sequence of a His-Pro-cGH sequence. The deduced sequence of the chymosin pro-peptide has been underlined and the deduced amino acid of cGH has been italicized. The cleavage site of enterokinase between (Lys31 - Asp32) and the peptide bond separating the pro-chymosin and the mature cGH (Phe84 - Ser85) are indicated with an arrow (↑). The poly-histidine site (His5-His10) and the enterokinase recognition site (Asp27 - Lys31) are also indicated.

Figure 3 is a schematic diagram of the His-Pro-cGH fusion construct. The enterokinase cleavage site (enterokinase cleavage) and pro-chymosin cleavage site (PRO Cleavage) are indicated with an arrow (↑).

Figure 4 illustrates the cleavage of purified His-Pro-cGH. Shown on the Western blot probed with an anti cGH antibodies are column purified His-Pro-cGH protein extracts from *E. coli* cells expressing the His-Pro-cGH fusion construct treated with enterokinase (Cut (EK)), mature chymosin at low pH (Cut (PRO)) and the control which was treated with PBS buffer (Uncut).

Figure 5 illustrates the cleavage of purified His-Pro-cGH. Shown on the Western blot probed with anti cGH antibodies are column purified His-Pro-cGH protein extracts from *E. coli* cells expressing the His-Pro-cGH fusion construct treated with mature chymosin at low pH (Cut (PRO)), treated with enterokinase (Cut (EK)), treated with gut extract from red turnip beetle (Cut (Red Turnip Gut)).

**Table 1: Activity measurements of bacterial extracts containing GST  
(Glutathion-S-transferase) and GST-Pro-Hirudin fusions.**

Sample	$\Delta$ Abs (405nm)/2min [Test 1]	$\Delta$ Abs (405nm)/2min [Test 2]
1 unit Thrombin	0.088	0.066
A: GST	0.087	0.082
B: GST pH 2.0	0.082	0.073
C: GST pH 2.0 +2 $\mu$ g chymosin	0.063	0.073
D: GST pH 4.5	0.087	0.086
E: GST pH 4.5 +2 $\mu$ g chymosin	0.087	0.087
A: GST-PRO-HIR	0.076	0.071
B: GST-PRO-HIR pH 2.0	0.072	0.064
C: GST-PRO-HIR pH 2.0 +2 $\mu$ g chymosin	0.066	0.070
D: GST-PRO-HIR pH 4.5	0.078	0.075
E: GST-PROHIR pH 4.5 +2 $\mu$ g chymosin	0.0002	0.0001
Hirudin 2 $\mu$ g	0.0001	0.0001



- 21 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANTS:

(A) NAME: SemBioSys Genetics Inc.  
(B) STREET: 2500 University Drive N.W.  
(C) CITY: Calgary  
(D) STATE: Alberta  
(E) COUNTRY: Canada  
(F) POSTAL CODE: T2N 1N4  
(G) TELEPHONE NO.: (403) 220-5161  
(H) TELEFAX NO.: (403) 220-0704

(A) NAME: van Rooijen, Gijs  
(B) STREET: 3223 Bearspaw Drive N.W.  
(C) CITY: Calgary  
(D) STATE: Alberta  
(E) COUNTRY: Canada  
(F) POSTAL CODE: T2L 1T1

(A) NAME: Alcantara, Joenel  
(B) STREET: 3 Castledale Place N.W.  
(C) CITY: Calgary  
(D) STATE: Alberta  
(E) COUNTRY: Canada  
(F) POSTAL CODE: T3J 1Y4

(A) NAME: Moloney, Maurice M.  
(B) STREET: 34 Edgebrook Cove, N.W.  
(C) CITY: Calgary  
(D) STATE: Alberta  
(E) COUNTRY: Canada  
(F) POSTAL CODE: T3A 5N5

(ii) TITLE OF INVENTION: Method for Cleavage of Fusion Proteins

(iii) NUMBER OF SEQUENCES: 4

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: BERESKIN & PARR  
(B) STREET: 40 King Street West  
(C) CITY: Toronto  
(D) STATE: Ontario  
(E) COUNTRY: Canada  
(F) ZIP: M5H 3Y2

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: PCT  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Gravelle, Micheline  
(B) REGISTRATION NUMBER: 40,261  
(C) REFERENCE/DOCKET NUMBER: 9369-54

- 22 -

## (ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (416) 364-7311  
 (B) TELEFAX: (416) 361-1398

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1096 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1032

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC	48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro	
1 5 10 15	
ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG	96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu	
20 25 30	
TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG	144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu	
35 40 45	
GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA	192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys	
50 55 60	
TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC	240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn	
65 70 75 80	
ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA	288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu	
85 90 95	
GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT	336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser	
100 105 110	
AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA	384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu	
115 120 125	
ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT	432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn	
130 135 140	
GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT	480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp	
145 150 155 160	
GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA	528
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu	
165 170 175	

- 23 -

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC	576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr	
180 185 190	
TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC	624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala	
195 200 205	
ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT	672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg	
210 215 220	
GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC GCT GAG ATC ACC	720
Gly Ser Pro Asn Ser Val Asp Ser Ser Gly Arg Ala Glu Ile Thr	
225 230 235 240	
AGG ATC CCT CTG TAC AAA GGC AAG TCT CTG AGG AAG GCG CTG AAG GAG	768
Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu	
245 250 255	
CAT GGG CTT CTG GAG GAC TTC CTG CAG AAA CAG CAG TAT GGC ATC AGC	816
His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser	
260 265 270	
AGC AAG TAC TCC GGC TTC GTC GTC TAT ACC GAC TGT ACC GAG TCC GGT	864
Ser Lys Tyr Ser Gly Phe Val Val Tyr Thr Asp Cys Thr Glu Ser Gly	
275 280 285	
CAG AAC CTC TGT CTC TGT GAG GGT TCC AAC GTC TGT GGT CAG GGT AAC	912
Gln Asn Leu Cys Leu Cys Glu Gly Ser Asn Val Cys Gly Gln Gly Asn	
290 295 300	
AAG TGT ATC CTC GGT TCC GAC GGT GAG AAG AAC CAG TGT GTC ACC GGT	960
Lys Cys Ile Leu Gly Ser Asp Gly Glu Lys Asn Gln Cys Val Thr Gly	
305 310 315 320	
GAG GGA ACC CCA AAG CCA CAG TCC CAC AAC GAC GGT GAC TTT GAG GAG	1008
Glu Gly Thr Pro Lys Pro Gln Ser His Asn Asp Gly Asp Phe Glu Glu	
325 330 335	
ATC CCA GAG GAG TAT CTC CAG TAA AGATCTAAGC TTGCTGCTGC TATCGAATTC	1062
Ile Pro Glu Glu Tyr Leu Gln *	
340	
CTGCAGCCCG GGGGATCCAC TAGTTCTAGA GCGG	1096

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 344 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Ser	Pro	Ile	Leu	Gly	Tyr	Trp	Lys	Ile	Lys	Gly	Leu	Val	Gln	Pro
1				5				10						15	
Thr	Arg	Leu	Leu	Leu	Glu	Tyr	Leu	Glu	Glu	Lys	Tyr	Glu	Glu	His	Leu
		20					25					30			
Tyr	Glu	Arg	Asp	Glu	Gly	Asp	Lys	Trp	Arg	Asn	Lys	Lys	Phe	Glu	Leu
	35					40					45				

- 24 -

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys  
     50                    55                    60  
 Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn  
     65                    70                    75                    80  
 Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu  
                     85                    90                    95  
 Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser  
                     100                    105                    110  
 Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu  
                     115                    120                    125  
 Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn  
                     130                    135                    140  
 Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp  
     145                    150                    155                    160  
 Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu  
                     165                    170                    175  
 Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr  
                     180                    185                    190  
 Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala  
                     195                    200                    205  
 Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg  
     210                    215                    220  
 Gly Ser Pro Asn Ser Arg Val Asp Ser Ser Gly Arg Ala Glu Ile Thr  
     225                    230                    235                    240  
 Arg Ile Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu  
                     245                    250                    255  
 His Gly Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser  
                     260                    265                    270  
 Ser Lys Tyr Ser Gly Phe Val Val Tyr Thr Asp Cys Thr Glu Ser Gly  
                     275                    280                    285  
 Gln Asn Leu Cys Leu Cys Glu Gly Ser Asn Val Cys Gly Gln Gly Asn  
     290                    295                    300  
 Lys Cys Ile Leu Gly Ser Asp Gly Glu Lys Asn Gln Cys Val Thr Gly  
     305                    310                    315                    320  
 Glu Gly Thr Pro Lys Pro Gln Ser His Asn Asp Gly Asp Phe Glu Glu  
                     325                    330                    335  
 Ile Pro Glu Glu Tyr Leu Gln \*  
                     340

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 819 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- 25 -

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..819

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT	48
Met Arg Gly Ser His His His His His His Gly Met Ala Ser Met Thr	
1 5 10 15	
GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT	96
Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr Asp Asp Asp Asp Lys Asp	
20 25 30	
CCG AGC TCG AGA TCT GCA GAA ATC GGA TCC GCT GAG ATC ACC AGG ATC	144
Pro Ser Ser Arg Ser Ala Glu Ile Gly Ser Ala Glu Ile Thr Arg Ile	
35 40 45	
CCT CTG TAC AAA GGC AAG TCT CTG AGG AAG GCG CTG AAG GAG CAT GGG	192
Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu His Gly	
50 55 60	
CTT CTG GAG GAC TTC CTG CAG AAA CAG CAG TAT GGC ATC AGC AGC AAG	240
Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser Ser Lys	
65 70 75 80	
TAC TCC GGC TTC TCA GAC AAC CAG CGG CTC TTC AAT AAT GCA GTC ATT	288
Tyr Ser Gly Phe Ser Asp Asn Gln Arg Leu Phe Asn Asn Ala Val Ile	
85 90 95	
CGT GTA CAA CAC CTG CAC CAG CTG GCT GCA AAA ATG ATT AAC GAC TTT	336
Arg Val Gln His Leu His Gln Leu Ala Ala Lys Met Ile Asn Asp Phe	
100 105 110	
GAG GAC AGC CTG TTG CCT GAG GAA CGC AGA CAG CTG AGT AAA ATC TTC	384
Glu Asp Ser Leu Leu Pro Glu Glu Arg Arg Gln Leu Ser Lys Ile Phe	
115 120 125	
CCT CTG TCT TTC TGC AAT TCT GAC TAC ATT GAG GCG CCT GCT GGA AAA	432
Pro Leu Ser Phe Cys Asn Ser Asp Tyr Ile Glu Ala Pro Ala Gly Lys	
130 135 140	
GAT GAA ACA CAG AAG AGC TCT ATG CTG AAG CTT CTT CGC ATC TCT TTT	480
Asp Glu Thr Gln Lys Ser Ser Met Leu Lys Leu Leu Arg Ile Ser Phe	
145 150 155 160	
CAC CTC ATT GAG TCC TGG GAG TTC CCA AGC CAG TCC CTG AGC GGA ACC	528
His Leu Ile Glu Ser Trp Glu Phe Pro Ser Gln Ser Leu Ser Gly Thr	
165 170 175	
GTC TCA AAC AGC CTG ACC GTA GGG AAC CCC AAC CAG CTC ACT GAG AAG	576
Val Ser Asn Ser Leu Thr Val Gly Asn Pro Asn Gln Leu Thr Glu Lys	
180 185 190	
CTG GCC GAC TTG AAA ATG GGC ATC AGT GTG CTC ATC CAG GCA TGT CTC	624
Leu Ala Asp Leu Lys Met Gly Ile Ser Val Leu Ile Gln Ala Cys Leu	
195 200 205	
GAT GGT CAA CCA AAC ATG GAT GAT AAC GAC TCC TTG CCG CTG CCT TTT	672
Asp Gly Gln Pro Asn Met Asp Asp Asn Asp Ser Leu Pro Leu Pro Phe	
210 215 220	
GAG GAC TTC TAC TTG ACC ATG GGG GAG AAC AAC CTC AGA GAG AGC TTT	720
Glu Asp Phe Tyr Leu Thr Met Gly Glu Asn Asn Leu Arg Glu Ser Phe	

- 26 -

225		230		235		240	
CGT CTG CTG GCT TGC TTC AAG AAG GAC ATG CAC AAA GTC GAG ACC TAC							768
Arg Leu Leu Ala Cys Phe Lys Lys Asp Met His Lys Val Glu Thr Tyr							
	245			250		255	
TTG AGG GTT GCA AAT TGC AGG AGA TCC CTG GAT TCC AAC TGC ACC CTG							816
Leu Arg Val Ala Asn Cys Arg Arg Ser Leu Asp Ser Asn Cys Thr Leu							
	260		265		270		
TAG							819
*							

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 273 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Arg Gly Ser His His His His His Gly Met Ala Ser Met Thr
 1           5           10           15
Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr Asp Asp Asp Asp Lys Asp
          20           25           30
Pro Ser Ser Arg Ser Ala Glu Ile Gly Ser Ala Glu Ile Thr Arg Ile
   35           40           45
Pro Leu Tyr Lys Gly Lys Ser Leu Arg Lys Ala Leu Lys Glu His Gly
   50           55           60
Leu Leu Glu Asp Phe Leu Gln Lys Gln Gln Tyr Gly Ile Ser Ser Lys
   65           70           75           80
Tyr Ser Gly Phe Ser Asp Asn Gln Arg Leu Phe Asn Asn Ala Val Ile
          85           90           95
Arg Val Gln His Leu His Gln Leu Ala Ala Lys Met Ile Asn Asp Phe
          100          105          110
Glu Asp Ser Leu Leu Pro Glu Glu Arg Arg Gln Leu Ser Lys Ile Phe
          115          120          125
Pro Leu Ser Phe Cys Asn Ser Asp Tyr Ile Glu Ala Pro Ala Gly Lys
          130          135          140
Asp Glu Thr Gln Lys Ser Ser Met Leu Lys Leu Leu Arg Ile Ser Phe
          145          150          155          160
His Leu Ile Glu Ser Trp Glu Phe Pro Ser Gln Ser Leu Ser Gly Thr
          165          170          175
Val Ser Asn Ser Leu Thr Val Gly Asn Pro Asn Gln Leu Thr Glu Lys
          180          185          190
Leu Ala Asp Leu Lys Met Gly Ile Ser Val Leu Ile Gln Ala Cys Leu
          195          200          205
Asp Gly Gln Pro Asn Met Asp Asp Asn Asp Ser Leu Pro Leu Pro Phe

```

- 27 -

210						215												220
Glu	Asp	Phe	Tyr	Leu	Thr	Met	Gly	Glu	Asn	Asn	Leu	Arg	Glu	Ser	Phe			
225					230					235					240			
Arg	Leu	Leu	Ala	Cys	Phe	Lys	Lys	Asp	Met	His	Lys	Val	Glu	Thr	Tyr			
				245					250					255				
Leu	Arg	Val	Ala	Asn	Cys	Arg	Arg	Ser	Leu	Asp	Ser	Asn	Cys	Thr	Leu			
			260					265					270					

\*

We Claim:

1. A method for the preparation of a recombinant polypeptide comprising
  - a) introducing into a host cell an expression vector comprising:
    - (1) a nucleic acid sequence capable of regulating transcription in a host cell,  
5 operatively linked to
    - (2) a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising (a) a nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen, linked in reading frame to (b) a nucleic acid sequence heterologous to the pro-peptide  
10 and encoding the recombinant polypeptide; operatively linked to
    - (3) a nucleic acid sequence encoding a termination region functional in said host cell,
  - b) growing the host cell to produce said fusion protein; and
  - c) altering the environment of the fusion protein so that the pro-peptide is  
15 cleaved from the fusion protein to release the recombinant polypeptide.
2. A method according to claim 1 wherein said pro-peptide is derived from a protease.
3. A method according to claim 1 wherein said pro-peptide is derived from an aspartic protease, a serine protease or a cysteine protease.
- 20 4. A method according to claim 1 wherein said pro-peptide is selected from the group comprising chymosin, trypsinogen, pepsin, HIV-1 protease, pepsinogen, cathepsin or yeast proteinase A.
5. A method according to claim 1 wherein the polypeptide is hirudin or carp growth hormone.
- 25 6. The method according to claim 1 wherein the chimeric nucleic acid sequence does not include a sequence encoding a mature form of the zymogen.
7. A method according to claim 1 wherein the altering the environment comprises altering the pH, altering the salt concentration or altering the temperature.
8. A method according to claim 7 wherein the altering the pH comprises  
30 altering the pH to a pH from about 2 to about 4.5.



- 29 -

9. A method according to claim 1 wherein the altering the environment takes place under *in vitro* conditions.
10. A method according to claim 1 wherein said altering the environment takes place under *in vivo* conditions.
- 5 11. A method according to claim 10 wherein the *in vivo* conditions are those prevalent in a tissue or bodily fluid of an animal.
12. A method according to claim 11 wherein the tissue or bodily fluid comprises the milk, blood, the stomach, the gut or the kidneys of said animal.
13. A method according to claim 1 wherein a mature form of an  
10 autocatalytically maturing zymogen is added in step (c) wherein said zymogen is homologous to the pro-peptide.
14. A method according to claim 1 wherein a mature form of an autocatalytically maturing zymogen is added in step (c) wherein said zymogen is heterologous to the pro-peptide.
- 15 15. The method according to claims 13 or 14 wherein the mature zymogen is added under *in vitro* conditions.
16. The method according to claims 13 or 14 wherein the mature zymogen is added under *in vivo* conditions.
17. The method according to claim 16 wherein said *in vivo* conditions are those  
20 prevalent in a tissue or bodily fluid of an animal.
18. The method according to claim 17 wherein the tissue or bodily fluid is a stomach, kidney, gut, blood or milk of said animal.
19. A method according to any one of claims 1 to 18 wherein said nucleic acid sequences are deoxyribonucleic acid (DNA) sequences.
- 25 20. A chimeric nucleic acid sequence encoding a fusion protein comprising (a) a nucleic acid sequence encoding a pro-peptide from an autocatalytically maturing zymogen

- 30 -

and (b) a nucleic acid sequence encoding a polypeptide that is heterologous to the pro-peptide.

21. A chimeric nucleic acid sequence according to claim 20 wherein the pro-peptide is derived from a protease.
- 5 22. A chimeric nucleic acid sequence according to claim 20 wherein the pro-peptide is derived from a serine protease, aspartic protease or a cysteine protease.
23. A chimeric nucleic acid sequence according to claim 20 wherein the pro-peptide is derived from chymosin, trypsinogen, pepsin, HIV-1 protease, pepsinogen, cathepsin or yeast proteinase A.
- 10 24. A chimeric nucleic acid sequence according to claim 20 wherein the polypeptide is hirudin or carp growth hormone.
25. A chimeric nucleic acid sequence according to claim 20 which does not include a sequence encoding a mature form of the zymogen.
26. A chimeric nucleic acid sequence according to any one of claims 20 to 25  
15 wherein said nucleic acid sequences are deoxyribonucleic acid (DNA) sequences.
27. A chimeric nucleic acid sequence according to claim 26 wherein the chimeric sequence is as shown in SEQ.ID.NO 1. or SEQ. ID. NO.2.
28. An expression vector comprising a chimeric nucleic acid sequence according to any one of claims 20 to 27 and a regulatory sequence suitable for expression in a host cell.
- 20 29. A transformed host cell containing an expression vector according to claim 28.
30. A transformed host cell containing an expression vector according to claim 28 wherein the host cell is a bacterial cell, a fungal cell, a plant cell or an animal cell.
31. A method of delivering a therapeutic or nutritional polypeptide to a  
25 human or animal comprising
- (a) providing a fusion protein comprising

- 31 -

- (i) a pro-peptide derived from an autocatalytically maturing enzyme, linked to
- (ii) a polypeptide that is heterologous to the pro-peptide and is a therapeutic or nutritional protein; and
- 5 (b) administering the fusion protein to the human or animal where the therapeutic or nutritional polypeptide is cleaved from the pro-peptide.
32. A method according to claim 31 wherein the mature form of an autocatalytically maturing zymogen is added in step (b).
- 10 33. A method according to claim 31 wherein said mature autocatalytically maturing zymogen is homologous to the pro-peptide.
34. A method according to claim 31 wherein said mature autocatalytically maturing zymogen is heterologous to the pro-peptide.
35. A method according to any one of claims 31 to 34 wherein said pro-peptide
- 15 is derived from a protease.
36. A method according to claim 35 wherein said protease is an aspartic protease, a serine protease or a cysteine protease.
37. A method according to claim 35 wherein said protease is chymosin, trypsinogen, pepsin, HIV-1 protease, pepsinogen, cathepsin or yeast proteinase A.
- 20 38. A method according to any one of claims 31 to 37 wherein the polypeptide is a vaccine, a peptide antibiotic, a cattle feed enzyme, a cytokine, a gastric lipase or a lactase.
39. A pharmaceutical composition comprising a fusion protein which comprises (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a
- 25 polypeptide that is heterologous to the pro-peptide, in admixture with a suitable diluent or carrier.
40. A food composition comprising a fusion protein which comprises a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is heterologous to the pro-peptide, in admixture with a suitable diluent or carrier.

- 32 -

41. A pharmaceutical composition comprising a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising (a) a first nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen and (b) a second nucleic acid sequence encoding a polypeptide that is heterologous to the pro-peptide.
42. A food composition comprising a chimeric nucleic acid sequence encoding a fusion protein, the chimeric nucleic acid sequence comprising (a) a first nucleic acid sequence encoding a pro-peptide derived from an autocatalytically maturing zymogen and (b) a second nucleic acid sequence encoding a polypeptide that is heterologous to the pro-peptide.
43. A composition according to claim 41 or 42 wherein the nucleic acid sequences are deoxyribonucleic acid (DNA) sequences.
44. A composition according to claim 41, 42 or 43 wherein said chimeric nucleic acid sequence does not include a sequence encoding a mature form of the zymogen.
45. A fusion protein comprising (a) a pro-peptide derived from an autocatalytically maturing zymogen and (b) a polypeptide that is heterologous to the pro-peptide.
46. A fusion protein according to claim 45 which does not include a mature form of the zymogen.
47. A use of a fusion protein comprising (i) a pro-peptide derived from an autocatalytically maturing enzyme, linked to (ii) a polypeptide that is heterologous to the pro-peptide and is a therapeutic or nutritional protein; to deliver a therapeutic or nutritional protein to a human or animal.

1/7  
**FIGURE 1**

Nucleotide sequence and the deduced amino acid sequence of a GST-Pro-Hirudin fusion.

1	ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC ACT CGA CTT CTT	60
1	M S P I L G Y W K I K G L V Q P T R L L	20
61	TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG TAT GAG CGC GAT GAA GGT GAT AAA	120
21	L E Y L E E K Y E E H L Y E R D E G D K	40
121	TGG CGA AAC AAA AAG TTT GAA TTG GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT	180
41	W R N K K F E L G L E F P N L P Y Y I D	60
181	GGT GAT GTT AAA TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC	240
61	G D V K L T Q S M A I I R Y I A D K H N	80
241	ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA GGA GCG GTT TTG	300
81	M L G G C P K E R A E I S M L E G A V L	100
301	GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT AAA GAC TTT GAA ACT CTC AAA GTT	360
101	D I R Y G V S R I A Y S K D F E T L K V	120
361	GAT TTT CTT AGC AAG CTA CCT GAA ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA	420
121	D F L S K L P E M L K M F E D R L C H K	140
421	ACA TAT TTA AAT GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT	480
141	T Y L N G D H V T H P D F M L Y D A L D	160
481	GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA GTT TGT TTT AAA	540
161	V V L Y M D P M C L D A F P K L V C F K	180
541	AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC TTG AAA TCC AGC AAG TAT ATA GCA	600
181	K R I E A I P Q I D K Y L K S S K Y I A	200
601	TGG CCT TTG CAG GGC TGG CAA GCC ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT	660
201	W P L Q G W Q A T F G G G D H P P K S D	220
661	CTG GTT CCG CGT GGA TCC CCG AAT TCC CGG GTC GAC TCG AGC GGC CGC GCT GAG ATC ACC	720
221	L V P R G S P N S R V D S S G R <u>A E I T</u>	240

2/7  
FIGURE 1 (Cont'd)

721	AGG ATC CCT CTG TAC AAA GGC AAG TCT CTG AGG AAG GCG CTG AAG GAG CAT GGG CTT CTG	780
241	<u>R I P L Y K G K S L R K A L K E H G L L</u>	260
781	GAG GAC TTC CTG CAG AAA CAG CAG TAT GGC ATC AGC AGC AAG TAC TCC GGC TTC GTC GTC	840
261	<u>E D F L O K O O Y G L S S K Y S G F V V</u>	280
841	TAT ACC GAC TGT ACC GAG TCC GGT CAG AAC CTC TGT CTC TGT GAG GGT TCC AAC GTC TGT	900
281	<u>Y T D C T E S G Q N L C L C E G S N V C</u>	300
901	GGT CAG GGT AAC AAG TGT ATC CTC GGT TCC GAC GGT GAG AAG AAC CAG TGT GTC ACC GGT	960
301	<u>G Q G N K C I L G S D G E K N Q C V T G</u>	320
961	GAG GGA ACC CCA AAG CCA CAG TCC CAC AAC GAC GGT GAC TTT GAG GAG ATC CCA GAG GAG	1020
321	<u>E G T P K P Q S H N D G D F E E I P E E</u>	340
1021	TAT CTC CAG TAA agatctaagcttgctgctgctatcgaattcctgcagccccgggggataccactagttctagagcgg	1096
341	<u>Y L Q *</u>	344

3/7  
FIGURE 2

Nucleotide sequence and the deduced amino acid sequence of a His-Pro-cGH fusion.

Poly Histidine Site

1 ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT GGT GGA CAG CAA 60  
1 M R G S H H H H H H G M A S M T G G Q Q 20

EK recognition site

61 ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCG AGC TCG AGA TCT GCA GAA ATC 120  
21 M G R D L T D D D D D K D P S S R S A E I 40

↑

121 GGA TCC GCT GAG ATC ACC AGG ATC CCT CTG TAC AAA GGC AAG TCT CTG AGG AAG GCG CTG 180  
41 G S A E I T R I P L Y K G K S L R K A L 60

181 AAG GAG CAT GGG CTT CTG GAG GAC TTC CTG CAG AAA CAG CAG TAT GGC ATC AGC AGC AAG 240  
61 K E H G L L E D F L Q K O O Y G I S S K 80

241 TAC TCC GGC TTC TCA GAC AAC CAG CGG CTC TTC AAT AAT GCA GTC ATT CGT GTA CAA CAC 300  
81 Y S G F S D N Q R L F N N A V I R V Q H 100

↑

301 CTG CAC CAG CTG GCT GCA AAA ATG ATT AAC GAC TTT GAG GAC AGC CTG TTG CCT GAG GAA 360  
101 L H Q L A A K M I N D F E D S L L P E E 120

361 CGC AGA CAG CTG AGT AAA ATC TTC CCT CTG TCT TTC TGC AAT TCT GAC TAC ATT GAG GCG 420  
121 R R Q L S K I F P L S F C N S D Y I E A 140

421 CCT GCT GGA AAA GAT GAA ACA CAG AAG AGC TCT ATG CTG AAG CTT CTT CGC ATC TCT TTT 480  
141 P A G K D E T Q K S S M L K L L R I S F 160

481 CAC CTC ATT GAG TCC TGG GAG TTC CCA AGC CAG TCC CTG AGC GGA ACC GTC TCA AAC AGC 540  
161 H L I E S W E F P S Q S L S G T V S N S 180

541 CTG ACC GTA GGG AAC CCC AAC CAG CTC ACT GAG AAG CTG GCC GAC TTG AAA ATG GGC ATC 600  
181 L T V G N P N Q L T E K L A D L K M G I 200

4/7FIGURE 2 (Cont'd)

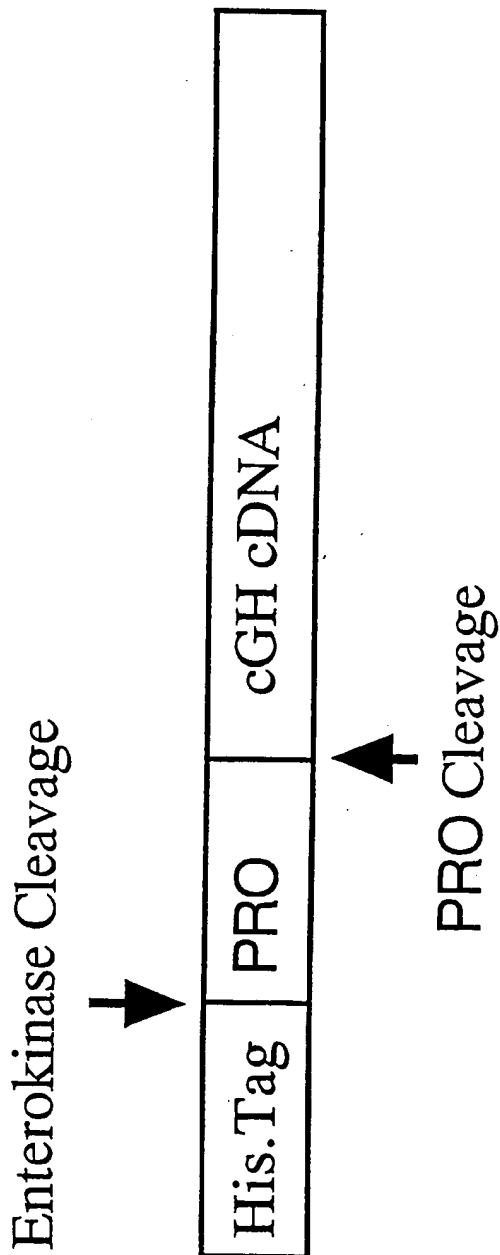
601 AGT GTG CTC ATC CAG GCA TGT CTC GAT GGT CAA CCA AAC ATG GAT GAT AAC GAC TCC TTG 660  
201 S V L I Q A C L D G Q P N M D D N D S L 220

661 CCG CTG CCT TTT GAG GAC TTC TAC TTG ACC ATG GGG GAG AAC AAC CTC AGA GAG AGC TTT 720  
221 P L P F E D F Y L T M G E N N L R E S F 240

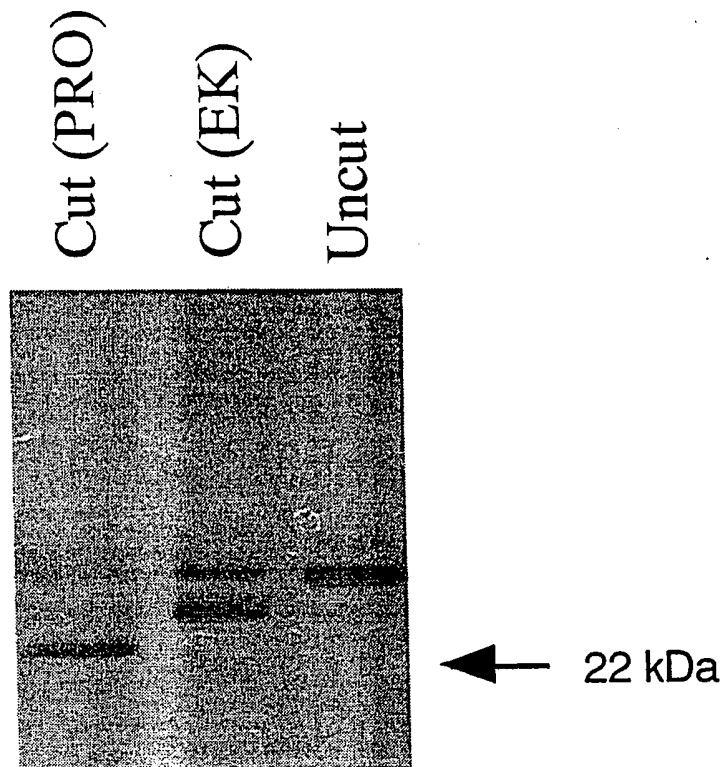
721 CGT CTG CTG GCT TGC TTC AAG AAG GAC ATG CAC AAA GTC GAG ACC TAC TTG AGG GTT GCA 780  
241 R L L A C F K K D M H K V E T Y L R V A 260

781 AAT TGC AGG AGA TCC CTG GAT TCC AAC TGC ACC CTG TAG  
261 N C R R S L D S N C T L \*

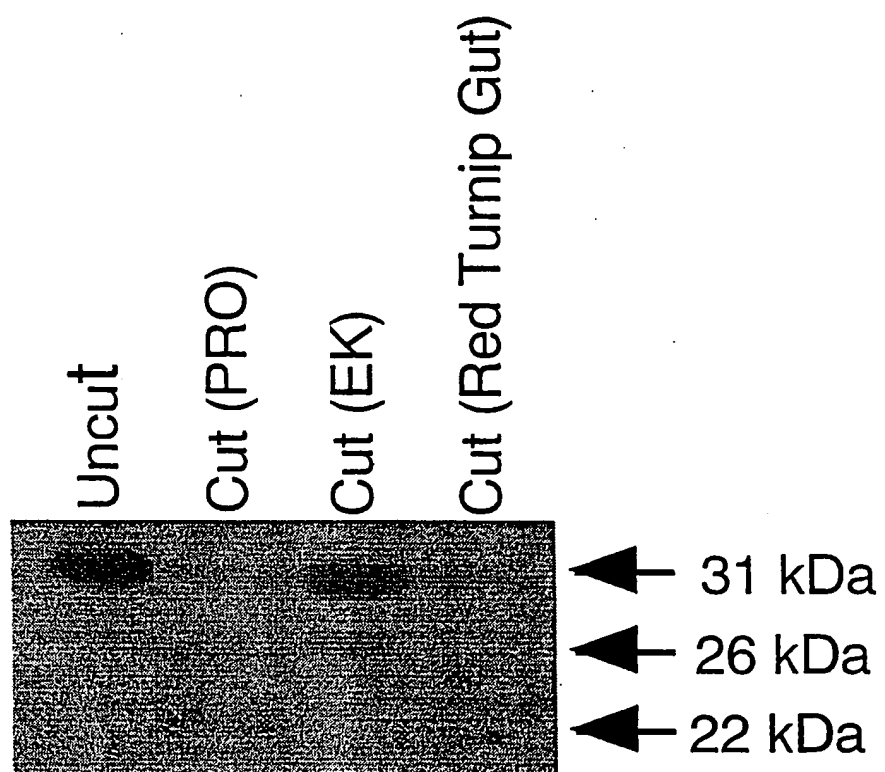


5/7FIGURE 3

6/7  
FIGURE 4



7/7  
FIGURE 5



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 98/00398

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/62 C12N15/15 C07K14/815 C07K14/61 C12N9/50  
C12N15/57 A61K38/58 A61K38/27 A23L1/305

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 21029 A (UNIV TECHNOLOGIES INT) 11 July 1996 cited in the application see page 4, line 13-29 see page 14, line 21-29 see page 20, line 1-10 see page 24, line 12-29 see page 29, line 3-14; example 8 see page 52, line 5-10 ---	1-47
X	EP 0 134 662 A (GENEX CORP) 20 March 1985 see the whole document --- -/--	1-47

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

21 July 1998

Date of mailing of the international search report

25.08.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Mateo Rosell, A.M.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CA 98/00398

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claim 31 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 98/00398

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9621029 A	11-07-1996	US 5650554 A	22-07-1997
		AU 4295096 A	24-07-1996
		CA 2208751 A	11-07-1996
		ZA 9510999 A	13-07-1996
EP 0134662 A	20-03-1985	AU 3036684 A	10-01-1985